Iranian Journal of Animal Science Research
Vol. 13, No. 2, Summer 2021, p. 313-319

نشریه پژوهشهای علوم دامی ایران
جلد ۱۳، شماره ۲، تابستان ۱۴۰۰، ص. ۳۱۹-۳۱۳

مقاله علمی- پژوهشی

شناسایی چند شکلی‌های تک‌نوکلئوتیدی (SNP) بر روی ترانسکریپتوم گاوهای هلشتاین آمریکا و کلیستانی پاکستان

مژگان قاسمی سیاب[۱]، شیدا ورکوهی[۲*]، محمدحسین بنابازی[۳]

تاریخ دریافت: ۱۳۹۸/۰۵/۰۷
تاریخ پذیرش: ۱۳۹۹/۰۳/۱۳

چکیده

مطالعه در سطح ترانسکریپتوم می‌تواند فاصله بین ژنوتیپ و فنوتیپ را پر نموده و به درک ساز وکارهای تبدیل توالی به عملکرد کمک نمایند. در این مطالعه به منظور کشف چند شکلی‌های تک‌نوکلئوتیدی (SNP) از داده‌های RNA-Seq دو جمعیت گاوهای هلشتاین آمریکا (*Bos taurus*) و کلیستانی پاکستان (*Bos indicus*) استفاده شد. کنترل کیفیت و ویرایش داده‌ها توسط نرم افزارهای FASTQC و Trimmomatic انجام شد. با همردیفی و مکان‌یابی خوانش‌های RNA-Seq بر روی ژنوم مرجع گاو با استفاده از نرم افزار TopHat2، ترانسکریپتوم تشکیل شد، سپس با استفاده از بسته نرم‌افزاری Samtools، آنالیز کشف SNP بر روی ترانسکریپتوم صورت گرفت که منجر به شناسایی ۵۰۱۸۳ و ۱۳۷۹۵۴ جایگاه SNP به ترتیب در گاوهای هلشتاین و کلیستانی شد، که ۱۵۳۰۸ جایگاه مشترک بود. ارتباط مستقیمی بین تعداد SNPهای یافت شده و طول کروموزوم‌ها مشاهده نشد. همچنین ۱۲ نوع SNP شناسایی شد که چهار نوع جایگزینی نوکلئوتیدی عامل و هشت نوع جایگزینی نوکلئوتیدی غیر عامل بودند. شایعترین SNPهای شناخته شده، از نوع جایگزینی نوکلئوتیدی عامل بودند، که ۷۰/۶٪ در نژاد کلیستانی و ۶۹/۶٪ در نژاد هلشتاین وجود داشت. درمطالعه حاضر ۱۲ نوع SNP شناسایی شد. چهار نوع از SNPها، از نوع جایگزینی نوکلئوتیدی عامل وهشت نوع جایگزینی نوکلئوتیدی غیر عامل بودند. شایعترین SNPهای شناخته شده، SNPهای از نوع جایگزینی نوکلئوتیدی عامل بودند، که به ترتیب ۷۰/۶٪ SNPها در نژاد کلیستانی و ۶۹/۶٪ SNPها در نژاد هلشتاین را شامل شد. نسبت جایگزینی نوکلئوتیدی عامل به غیر عامل (Ts/Tv) SNPها در نژاد کلیستانی برابر با ۲/۴ و در نژاد هلشتاین برابر با ۲/۳ بود.

واژه‌های کلیدی: ترانسکریپتوم، جایگزینی نوکلئوتیدی، چند شکلی تک نوکلئوتیدی، داده‌های RNA-Seq

مقدمه

چند شکلی‌های تک نوکلئوتیدی (SNP)[۵] امروزه به مهمترین و کارامدترین ابزار بهنژادی تبدیل شده‌اند. مطالعات پویش کل ژنوم روش مهمی در اصلاح نژد دام است که مبتنی بر تعیین ژنوتیپ تعداد بسیار زیادی از این جایگاه‌ها است. بدین منظور آرایه‌های با تراکم بالا از این نشانگر در صنعت گاو شیری ارائه شده‌اند. از آنجایی که ترانسکریپتوم تنها حاصل رونویسی بخش‌هایی از ژنوم است که رمزگر بوده و در نهایت به یک فراورده پروتئینی ترجمه و بیان می‌شود، احتمالا SNPهای موجود در این مناطق بهتر می‌توانند ارزش اصلاحی ژنومی را برآورد کنند (۹). ترانسکریپتوم می‌تواند به درک بهتر ساز و کارهای حد واسط میان ژنوتیپ و فنوتیپ کمک نماید و صحت و دقت پیش‌بینی های ژنومی را که امروزه مهمترین ابزار ارزیابی و انتخاب برای به نژادگران دامی می‌باشد، بهبود بخشیده و پیشرفت ژنتیکی را سریعتر نماید (۷). جامعیت ترانسکریپتوم به حدی است که محققین معتقدند می‌توان به جای ژن، ترانسکریپت (رونوشت)، یعنی قطعه ای از RNA که از روی DNA کد می‌شود را به عنوان واحد

۱- دانش آموخته کارشناسی ارشد، گروه علوم دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه رازی،کرمانشاه، ایران.
۲- استادیار، گروه علوم دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه رازی، کرمانشاه، ایران.
۳- استادیار، بخش پژوهشهای بیوتکنولوژی، موسسه تحقیقات علوم دامی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی،کرج، ایران.
(*- نویسنده مسئول: (Email: S.varkoohi@gmail.com
DOI:10.22067/ijasr.v13i2.82194
۵- *SINGLE NUCLEOTIDE POLYMORPHISM*

بنیادی ژنوم و واحد اســاســی توارث پیشــنهاد کرد (۳). داده‌های ترانسکریپتومی به اشکال مختلف تولید می‌گردند که آخرین نسل آنها حاصـل توالی‌یابی RNA (RNA-Seq) می‌باشـند. داده‌های RNA-Seq نوعی از داده‌های ترانسکریپتومی پربرونداد و بسیار دقیق و کمی هســتند. این داده‌ها علاوه بر توالی نوکلئوتیدی کل ترانســکریپتوم، همزمان میزان رونویسی در هر یک از نواحی آن را به صورت بسیار کمی و دقیق بدست می‌آورند (۳ و ۱۰).

طی پژوهشی با استفاده از داده‌های RNA-Seq بیش از ۳۳۰۰۰ نشانگر SNP در ترانسکریپتوم شیر گاو شناسایی شد (۲). همچنین طی پژوهش دیگری در دو بافت کبد و کلیه در ۵ نژاد بز به ترتیب ۶۸۵۹۷ و ۷۲۰۴۷ نشانگر SNP شناسایی شد (۷). طی تحقیقی روی دو بافت عضله و خون در شش اسب مسابقه نژاد تروبرد پیش و پس از مسابقه، تعداد ۱۸۳۹۷۳ جایگاه SNP کشف شد (۶). همچنین تعداد ۵۶۲۳۱ نشانگر SNP در پوست بز کشمیری چینی مشاهده شد (۱۱).

جایگزینی نوکلئوتیدی عامل یک جهش نقطه‌ای اســت که یک نوکلئوتید پورین به پورین دیگر ($A \leftrightarrow G$) یا یک نوکلئوتید پریمیدین به پریمیدین دیگر ($T \leftrightarrow C$) تبدیل می‌شــود، در حالی که جایگزینی نوکلئوتیدی غیر عامل باعث تغییر از نوکلئوتید پورین به پریمیدین یا برعکس ($A \leftrightarrow T$، $A \leftrightarrow C$، $G \leftrightarrow T$، $G \leftrightarrow C$) می‌شــود. اگر جهش‌ها روند تصــادفی داشــته باشــند، بایســتی نســبت جایگزینی نوکلئوتیدی عامل و غیر عامل یک به دو باشــد. طی تحقیقی بر روی اسـب کاسـپین، شـمار جایگزینی نوکلئوتیدی عامل برابر ۱۱۵۵۴۱۷ و شـمار جایگزینی نوکلئوتیدی غیر عامل برابر ۵۱۲۹۸۶ بود که نسـبت (Ts/Tv) ۲/۲۵ برآورد شد (۱). هدف از مطالعه اخیر کشف و شناسایی SNPهای موجود بر روی توالی ترانســکریپتوم نمونه‌ای از جمعیت گاوهای هلشتاین آمریکا و گاوهای کلیستانی پاکستان می‌باشد.

مواد و روشها

در این مطالعه از داده‌های RNA-Seq مربوط به ادغام ۴۰ نمونه خون گاو هلشتاین در دانشگاه ویسکانسین آمریکا (*Bos taurus*) و ۴۵ گاو ماده کلیســتانی (*Bos indicus*) در ایالت پنجاب پاکســتان استفاده شد. این داده‌ها در مطالعه هوانگ و همکاران (۴) تولید شد.

داده‌های مربوطه برای نژاد کلیســتانی[1] و نژاد هلشــتاین[2] از بخش آرشیو خوانش‌های کوتاه[3] (SRA) بانک اطلاعاتی مرکز ملی اطلاعات بیوتکنولوژی[4] (NCBI) دریافت شــد. داده‌های مورد آنالیز شــامل ۲۱۰۷۸۴۷۷ و ۲۰۹۴۰۰۶۳ خوانش کوتاه جفتی به طول ۷۵ جفت باز به ترتیب برای نژادهای هلشــتاین و کلیســتانی بود. تمام آنالیزها در محیط سیستم عامل لینوکس اوبونتو (Ubuntu) نسخه ۱۶٫۰۴ انجام شــدند. فرمت sra به خاطر حجم کم صــرفا برای بارگذاری بر روی بانکهای اطلاعاتی مناســب اســت و برای آنالیز ابتدا باید به فرمتهای دیگر تبدیل گردد که یکی از مرسـومترین آنها فرمت fastq اسـت که این تبدیل با دســتور fastq-dump از مجموع نرم افزار sratoolkit نسخه Ubuntu64 2.5.2 انجام شد. کنترل کیفیت و ویرایش داده‌ها توسط نرم افزارهای FASTQC و Trimmomatic انجام شد که بر اسـاس حذف آداپتور و نیز حذف و یا ویرایش نوکلئوتیدهای بی‌کیفیت از خوانش‌ها با آســتانه حداقل طول نهایی ۵۰ جفت باز برای خوانش‌های ویرایش شــده تنظیم شــد. با همردیفی و مکان‌یابی خوانش‌های RNA-Seq بر روی ژنوم مرجع گاو با استفاده از نرم افزار TopHat2، ترانسـکریپتوم تشـکیل شـد، سـپس با اسـتفاده از بسـته نرم‌افزاری Samtools، آنالیز کشــف SNP بر روی ترانســکریپتوم صورت گرفت.

نتیجه و بحث

ویرایش داده‌ها: با توجه به کیفیت نسبتا مطلوب داده‌ها، تعداد خوانش‌های بی‌کیفیت کنار گذاشــته شــده برای دو نژاد، تقریبا برابر و نسبت به تعداد کل خوانش‌ها کم بود. برای مثال، در نژاد کلیستانی از مجموع ۲۰۹۴۰۰۶۳ خوانش اولیه بعد از ویرایش تعداد ۱۹۳۷۹۴۸۷ خوانش، کیفیت مناســب داشــته و باقی ماندند. به عبارت دیگر ۱۵۶۰۵۷۶ خوانش (حدود ۷/۵ درصــد از کل خوانش‌ها) از آنالیز کنار گذاشته شد.

از ۱۹۳۷۹۴۸۷ خوانش پیشــرو و پســرو ویرایش شــده برای نژاد کلیســتانی به ترتیب ۱۵۷۵۴۵۹۳ (۸۱/۳٪) و ۱۵۴۸۴۸۵۵ (۷۹/۹٪) خوانش به درستی مکان‌یابی شدند. بیش از نه درصد یعنی معادل بیش از یک و نیم میلیون خوانش با چند مکان همردیف شــدند و به عبارت دیگر همردیفی چندگانه (Multiple alignments) داشــتند. از این میان بیش از ۶۶۰۰۰ خوانش با بیش از ۲۰ مکان ژنومی همردیف شدند. نرخ مکان‌یابی کل (Overall read mapping rate) برای نژاد کلیسـتانی ۸۰/۶٪ بدسـت آمد. علاوه بر این، تعداد ۱۴۰۱۰۰۳۱ جفت خوانش دو ســویه همردیف گردیدند که از این تعداد ۹/۴٪ (معادل ۱۳۱۵۱۰۴ جفت خوانش) همردیفی چندگانه و ۳٪ (معادل ۴۲۰۹۷۳ جفت خوانش) همردیفی ناجور (Discordant alignments) داشـتند. به عبارت دیگر، نرخ همردیفی جفتی جور شـده (Concordant pair alignment rate) ۷۰٪ بدست آمد.

نرخ مکان یابی برای خوانش های پیش رو و پس رو در نژاد

1- http://www.ncbi.nlm.nih.gov/sra/SRS454433

2- http://www.ncbi.nlm.nih.gov/sra/SRS454432

3- *SEQUENCE READ ARCHIVE*

4- *NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION*

هلشتاین بترتیب ۶۶/۳ و ۵۵/۴٪ به دست آمد. همچنین نرخ همردیفی چندگانه به ترتیب ۷/۳ و ۷/۱٪ و تعداد خوانش‌های با همردیفی بر روی بیش از بیست موقعیت ژنومی در حدود شصت هزار بود. نرخ همردیفی کل ۶۰/۸٪ و نرخ همردیفی جور شده ۵۱/۳٪ بدست آمد. در مجموع نرخ همردیفی در حد مطلوب ارزیابی شد.

کشف SNPها: با همردیفی خوانش‌های کوتاه بر روی ژنوم مرجع RNA-Seq و سپس همگذاری آنها، بر روی ترانسکریپتوم نمونه جمعیت گاو هلشتاین آمریکا و کلیستانی پاکستان به ترتیب ۵۰۱۸۳ و ۱۳۷۹۵۴ نشانگر SNP شناسایی شد. این دو فهرست ۱۵۳۰۸ جایگاه SNP مشترک با هم داشتند. تعداد SNP کشف شده برای نژاد کلیستانی تقریبا سه برابر نژاد هلشتاین بود. این امر احتمالا بواسطه این است که برای همردیفی هر دو نژاد، که یکی از آنها یعنی هلشتاین متعلق به گونه گاوهای *Bos taurus* و دیگری یعنی کلیستانی یک گاو زبو (*Zebu*) و متعلق به گونه *Bos indicus* است، از یک ژنوم مرجع یکسان با منشاء گاو هرفورد (که آن هم از دسته گاوهای *Bos taurus* است) استفاده شد. علاوه بر این در همردیفی از تنظیمات سخت گیرانه (Stringent) برنامه tophat2 استفاده نشد به گونه‌ای که ممکن است با وجود تعداد زیادی عدم انطباق (Mismatch) بین نوکلئوتیدهای واقع بر ترانسکریپتوم گاو کلیستانی و ژنوم مرجع، باز هم همردیفی بطور موفق انجام شود. از اینرو، بعدا در آنالیز کشف SNP، همه این عدم انطباق‌ها به عنوان SNP در نظر گرفته می‌شوند. علاوه بر این، تنظیمات گفته شده منجر به بالاتر رفتن نسبی نرخ همردیفی و مکان یابی نیز می‌شود. بخشی از SNPهای اضافی کشف شده بر روی توالی ترانسکریپتوم گاو کلیستانی نیز ممکن است ناشی از حدود ۲۰٪ نرخ بالاتر همردیفی و مکان یابی در کلیستانی نسبت به هلشتاین (۷۰/۱ در مقابل ۵۱/۳ درصد) باشد (جدول ۱).

جدول ۱– تعداد SNP کشف شده در هر یک از دو نژاد هلشتاین و کلیستانی به تفکیک هر کروموزوم

Table 1- The number of discovered SNP in two Holstein and Cholistani breed based on chromosome

کروموزوم Chromosome	طول کروموزوم Chromosome length (bp)	SNPهای کشف شده در گاوهای کلیستانی discovered SNP in Cholistani cows	SNPهای کشف شده در گاوهای هلشتاین discovered SNP in Holstein cows
1	158337067	5590	2079
2	137060424	5764	2277
3	121430405	6865	2741
4	120829699	4356	1837
5	121191424	7854	2872
6	119458736	3743	1572
7	112638659	7287	2423
8	113384836	3883	1550
9	105708250	2798	1075
10	104305016	5264	2110
11	107310763	7191	2511
12	91163125	3057	1228
13	84240350	5388	2035
14	84648390	2962	974
15	85296676	4164	1583
16	81724687	4474	1501
17	75158596	4804	1586
18	66004023	7879	2750
19	64057457	8426	2914
20	72042655	2085	760
21	71599096	3684	1279
22	61435874	4052	1257
23	52530062	5263	2457
24	62714930	2225	782
25	42904170	5694	1798
26	51681464	2735	887
27	45407902	1473	524
28	46312546	1982	720
29	51505224	3809	1120
X	148823899	2783	836
ژنوم میتوکندریایی mitochondrial genome	16338	37	0
توالی‌های غیر کروموزومی Non-chromosomal sequences	9499556	383	145
کل total	2670422299	137954	50183

نسبت تعداد SNP به طول کروموزوم‌های مختلف نشان می‌دهد که ارتباط مستقیمی بین تعداد SNP های یافت شده و طول

کروموزوم‌ها وجود ندارد (جدول ۱). می‌توان نتیجه گرفت که رونویسی در تمام طول ژنوم با یک توزیع همگن و با پو ش شی یک سان صورت نمی‌گیرد. به ع بارت دیگر، احت مالا برخی نواحی حا مل ژن های کاندیدای بیشـــتر و یا ژن‌هایی هســـتند که برای جمعیت فوق اهمیت بیشـــتری داشـــته و رونویســـی (بیان) در آن مناطق شـــدیدتر و با عمق بی شتری صورت گرفته ا ست و در نتیجه این نواحی سهم بی شتری از کل تراذ سکریپتوم ا سمبل شده دارند. بالطبع SNPهای موجود در این نواحی نیز فراوانی بالاتر دا شته و بعد از فیلتر همچنان در لی ست باقی مانده‌اند.

انواع SNP: در مطالعه حاضر ۱۲ نوع SNP شناسایی شد. چهار نوع از SNP ها، از نوع جایگزینی نوکلئوتیدی عا مل (transition) وهشـــت نوع از SNP ها، از نوع جایگزینی نوکلئوتیدی غیر عا مل (transversion) بودند. شایعترین SNPهای شناخته شده، SNPهای از نوع جایگزینی نوکلئوتیدی عا مل بود ند، که به ترتیب ۷۰/۶٪ SNPها در نژاد کلیستانی و ۶۹/۶٪ SNPها در نژاد هلشتاین را شامل شـــد. نســـبت جایگزینی نوکلئوتیدی عامل به جایگزینی غیر عامل (Ts/Tv) برای SNP ها در نژاد کلیســـتانی برابر با ۲/۴ و در نژاد هلشـــتاین برابر با ۲/۳ بود (جدول ۲)، که این نتایج با نتایج مطالعه عارف نژاد و همکاران (۱) بر روی ا سب کا سپین و Jiang و همکاران (۵) روی ژنوم گاو مطابقت دارد. احتمال اینکه جایگزینی نوکلئوتیدی غیرعامل نســـبت به جایگزینی نوکلئوتیدی عامل توالی اســـیدآمینه پروتئین را تغییر دهد، بیشـــتر اســـت. هر دو نوع جایگزینی می‌توانند منجر به جایگزینی اســـیدآمینه شـــوند، اما تفاوت‌های بیوشـــیمیایی محصـــولات پروتئینی مرتبط با آن ها برای جایگزینی نوکلئوتیدی غیرعامل بیشتر است. از آنجا که جایگزینی نوکلئوتیدی غیرعامل علت تغییرات بزرگ در شـکل رشـته DNA اسـت، تاثیر بیشـتری بر روی اتصـــال فاکتور های رونویســـی و بنابراین بیان ژن دارد. اگر چه تفاوت‌های م شاهده شده کوچک ه ستند، اما این یافته‌ها یک ویژگی جدید و اســاســی تغییرات تنظیمی را نشــان می‌دهند. انحرافات ایجاد شده در نسبت (Ts/Tv) نشان دهنده انتخاب تکاملی ژن است (۱).

جدول ۲- انواع SNP در هر یک از دو نژاد هلشتاین و کلیستانی
Table 2- SNP types in two Holstein and Cholistani breed

	کلیستانی Cholistani		هلشتاین Holstein	
انواع SNP SNP types	1	2	1	2
A/C	5564	4	1867	3/7
C/A	5105	3/7	2399	4/8
A/T	3385	2/5	1342	2/7
T/A	3328	2/4	1344	2/7
C/G	6131	4/5	1996	4
G/C	6227	4/5	2085	4/1
G/T	5249	3/8	2301	4/6
T/G	5584	4	1893	3/8
C/T	23602	17/1	8795	17/5
T/C	24954	18/1	8722	17/4
A/G	24862	18	8788	17/5
G/A	23963	17/4	8651	17/2
کل Total	37954	100	50183	100

[1] Number of SNP type
[2] Percent of SNP type

نتیجه‌گیری کلی

در مجموع نتایج مطالعه حاضـر دقت و صـحت بالای شـناسـایی SNP های بدست آمده و نیز کارایی داده‌های RNA-Seq را در کشف SNP تایید می‌کند و علت تفاوت تعداد SNP کشـف شـده در مطالعه حا ضر و مطالعات ذکر شده احتمالا نا شی از عوامل تاثیرگذار متعددی از جم له تفاوت گو نه، نژاد، با فت، تعداد نمو نه مورد م طالعه، عمق توالی‌یابی و ... اســـت. تفاوت بیان بین دو آلل در یک موقعیت تک نوکلئوتیدی، زمینه ســاز تنوع فنوتیپی شــده و احتمالا بخش زیادی از واریانس بین این دو زیر گونه گاو به ویژه از نظر تنوع، حســاســیت به بیماری، تحمل تنش های محیطی اعم از تنش های زیســـتی و غیر زیستی، صفات تولیدی و عملکردی نژاد در مقابله با انگل‌ها و شرایط محیطی مختلف را توجیه می‌کند. انتخاب نژاد کلیستانی به عنوان یک نژاد گاو گوشـتی و زبو در کنار هلشـتاین به عنوان یک نژاد شـیری و متمایز برای آن است که دریابیم فهرست‌های SNP آنها چقدر متفاوت ه ستند. به این ترتیب وابـستگی فهر ست‌های م ستقل آنها به نوع نژاد

مشخص می‌گردد. در نهایت می‌توان در کنار فهرست‌های مجزا، یک فهرست مشترک ارائه داد که در هر دو نژاد کارایی داشته باشد.

این مطالعه فهرستی از تعداد قابل ملاحظه ای SNP بر ترانسکریپتوم جمعیتی از گاوهای هلشتاین و کلیستانی ارایه می‌دهد که بسیاری از آنها در مقایسه با SNP های موجود در آرایه‌های مرسوم، جدید هستند. مطالعه حاضر سر آغازی بر ورود بر عصر جدیدی از ارزیابی و انتخاب در حیوانات اهلی است که می‌تواند با بهره گیری از سطوح مختلف داده‌های اومیکس ادامه یابد و روند پیشرفت ژنتیکی را در جمعیتهای دامی بهبود بخشد.

منابع

1. Arefnezhad, B., H. Kohram, M. Moradi Shahre Babak, M. Shakeri, Y. Dong, X. Zhang, W. Wang, and G. H. Hosseini Salekdeh. 2015. Genetic Variant Detection of Caspian Horse Using High-throughput Sequecing Technology (in Persian). Journal of Agriculture Biotechnology, 4:101-116. (In Persian).
2. Cánovas, A., G. Rincon, A. Islas-Trejo, S. Wickramasinghe, and J. F. Medrano. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mammalian genome, 21:592-598.
3. Flintoft, L. 2008. Transcriptomics: Digging deep with RNA-Seq. Nature Reviews Genetics, 9(8):568- 568.
4. Huang, W., A. Nadeem, B. Zhang, M. Babar, M. Soller, and H. Khatib. 2012. Characterization and Comparison of the Leukocyte Transcriptomes of Three Cattle Breeds. Plos one, 7(1):e30244.
5. Jiang, Z., X. L. Wu, M. Zhang, L. M. Jennifer, and R. W. Wright. 2008. Thw complementary neighborhood patterns and methylation to mutation Likelihood structure of 15, 110 Single Nucleotide Polymorphisms in the bovine genome. Genetics, 180 (1): 639-647.
6. Park, K. D., J. Park, J. Ko, B. C. Kim, H. S. Kim, K. Ahn, K. T. Do, H. Choi, H. M. Kim, S. Song, S. Lee, S. Jho, , H. S. Kong, Y. M. Yang, B. H. Jhun, C. Kim, T. H. Kim, S. Hwang, J. Bhak, H. K. Lee, and B. W. Cho. 2012. Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. BMC Genomics, 13:473.
7. Pennisi, E. 2012. ENCODE Project Writes Eulogy for Junk DNA. Science, 337(6099):1159 - 1161.
8. Sharma, U., P. Banerjee, J. Joshi, and R. K. Vijh. 2012. Ubiquitous Expression of Genes in tissues of Goat (*Capra hircus*) Using RNA-seq. International Journal of Animal and Veterinary Advances, 4(4):292 - 302.
9. Meuwissen, T, and M. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics, 185:623 - 631.
10. Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, 10:57 - 63.
11. Wang, L., Y. Zhang, M. Zhao, R. Wang, R. Su, and J. Li. 2015. SNP Discovery from Transcriptome of Cashmere goat skin. Asian-Australasian Journal of Animal Science, 28(9):1235 - 1243.

Iranian Journal of Animal Science Research
Vol. 13, No. 2, Summer 2021, p. 313-319

نشریه پژوهشهای علوم دامی ایران
جلد ۱۳، شماره ۲، تابستان ۱۴۰۰، ص. ۳۱۹-۳۱۳

Single Nucleotide Polymorphism (SNP) Discovery on Transcriptomes of American Holstein and Pakistanian Cholistani Cows

Mojgan Ghasemi-Siab[1], Sheida Varkoohi*[2] and Mohammad Hossein Banabazi[3]
Submitted: 29-07-2019
Accepted: 02-06-2020

Introduction (SNPs) are single nucleotide base variations, caused by transitions (C/T or G/A) or transversions (C/G, C/A, or T/A, T/G), in the same position between individual genomic DNA sequences. Single nucleotide polymorphisms have been applied as important molecular markers in genetics and breeding studies. About 40% of the Single nucleotide polymorphisms in the genes cause a change in an amino acid. The rapid advance of next generation sequencing provides a high-throughput means of SNP discovery. Transcriptome study can fill the gap between genotype and phenotype and help understanding the mechanisms from sequence to function. RNA sequencing (RNA-Seq) is a next generating sequencing based technology for studying of whole transcriptome and gene expression. It simultaneously enables study of transcriptomics sequences and very accurate quantitative gene expression (digital expression). Hence, these data are very suitable for high-throughput study of expression level of all transcribed genes and their SNPs (Single Nucleotide Polymorphism. Recently, RNA-Seq has also been used as an efficient and cost-effective method to systematically identify SNPs in transcribed regions in different species. A transcriptomics-based sequencing approach offers a cheaper alternative to identify a large number of polymorphisms and possibly to discover causative variants.

Materials and Methods In this study, RNA-Seq data were used to SNP discovery in American Holstein (*Bos taurus*) and Pakistanian Cholistani (*Bos indicus*) cows. RNA-Seq data of 21,078,477 and 20940063 paired end reads with 75 bp length resulted from pooling of whole blood samples of 40 Holstein cows at the University of Wisconsin, Dairy Cattle Center, USA, and 45 Cholistani cows at Gujait Peer Farm, Bahawalpur, Punjab, Pakistan, respectively, obtained from SRA database in NCBI for Holstein cows (http://www.ncbi.nlm.nih.gov/sra/SRX317197) and Cholistani cows http://www.ncbi.nlm.nih.gov/sra/SRS454433). MRNA sequencing was run on Illumina Genome Analyzer IIx (Illumina Inc., San Diego, CA). Data were converted from Sra format to Fastq format by fastq-dump command from Ubuntu linux version of Sratoolkit 2.5.4-1. Data quality control was checked by FastQC (v0.11.3) likewise trimmed for linked adaptors and bad quality reads by Trimmomatic 0.33 Adaptors were considered according to sequencing instrument as default (TruSeq2-PE.fa) and the minimum read length was set at 50 bp. Trimmed reads were aligned on UMD3.1 reference genome (release 81) based on annotation data by Tophat2, which applies Bowtie2 as the aligner. The transcriptome was assembled by TopHat2 software in two cow's population by aligning and mapping the RNA-Seq reads on bovine reference genome. The SNPs were discovered by Samtools software.

Results and Discussion After data editing, the removed and low quality reads in both breeds were almost equal and relatively low. The length of whole transcriptome assembled, for example 52798651 bases in Holstein, indicates around 2% of the whole genome (around 2.6 Mbp) expressed as mRNA. In Cholistani cows, read mapping rate for forward and reverse reads were 81.3 and 79.9%, respectively, and multiple alignments rate was about 9.4%. Overall read mapping was 80.6% and concordant pair alignment was 70.1%. In Holstein cows, read mapping rate for forward and reverse reads were 66.3 and 55.4%, respectively, and multiple alignments rate was about 7.2%,. Overall read mapping was 60.8% and concordant pair alignment was 51.3%. Results show that 50183 and 137954 SNPs were discovered on the assembled transcriptome of Holstein and Cholistani cow's samples, respectively, and 15308 SNPs were common in both breeds. No direct relation was found between the number of discovered SNPs and the chromosome length. Also 12 SNP types were identified including 4 transition and 8 transversion. The most commonly discovered SNP were transition, which were 70.6% in Cholistani and 69.6% in Holstein cows. The ratio of transition to transversion SNP (Ts / Tv) was 2.4 and 2.3 in Cholistani and Holstein cows, respectively. The number of discovered SNPs in Cholistani cows were approximately three times higher than Holstein cows. Because, for the alignment of both species used a same reference

1. Master Graduated, Department of Animal Science, College of Agriculture & Natural Resources, Razi University, Kermanshah, Iran.
2. Assistant Professor, Department of Animal Science, College of Agriculture & Natural Resources, Razi University, Kermanshah, Iran.
3. Assistant Professor, Department of Biotechnology, Animal Science Research Institute of IRAN (ASRI), Agricultural Research, Education & Extension Organization (AREEO), Karaj, Iran.
(*- Corresponding Author Email: s.varkoohi@gmail.com)
DOI:10.22067/ijasr.v13i2.82194

genome with Herford origin.

Conclusion the expression difference between two alleles in a single-nucleotide position causes phenotype diversity and probably explains the large part of variances between these two bovine subspecies, especially in diversity, susceptibility to disease and parasites, tolerating environmental stress such as biological and non-biological stresses in different environmental conditions. While, differential gene expression analysis or even allelic specific expression in gene level may not be able to explain phenotype diversity.

Key words: Chromosome length, Nucleotide replacement, RNA-Seq data.